



Regularized Gradient Algorithm for Non-Negative Independent Component Analysis

Wendyam S. B. Ouedraogo, Mériem Jaidane, Antoine Souloumiac, Christian Jutten

► To cite this version:

Wendyam S. B. Ouedraogo, Mériem Jaidane, Antoine Souloumiac, Christian Jutten. Regularized Gradient Algorithm for Non-Negative Independent Component Analysis. ICASSP 2011 - IEEE International Conference on Acoustics, Speech and Signal Processing, May 2011, Prague, Czech Republic. pp.2524-2527. hal-00596772

HAL Id: hal-00596772

<https://hal.science/hal-00596772>

Submitted on 30 May 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

REGULARIZED GRADIENT ALGORITHM FOR NON-NEGATIVE INDEPENDENT COMPONENT ANALYSIS

W. S. B. Ouedraogo^{1,2,3}

M. Jaidane²

A. Souloumiac¹

C. Jutten³

¹CEA, LIST, Laboratoire d'Outils pour l'Analyse de Données, Gif-sur-Yvette, F-91191, France ;

²Unité Signaux et Systèmes, National School of Engineers of Tunis, BP 37, 1002 Tunis, Tunisia

³GIPSA-lab, UMR 5216 CNRS, University of Grenoble, 961 rue de la Houille Blanche BP 46
F-38402 Grenoble Cedex, France

ABSTRACT

Independent Component Analysis (ICA) is a well-known technique for solving blind source separation (BSS) problem. However “classical” ICA algorithms seem not suited for non-negative sources. This paper proposes a gradient descent approach for solving the Non-Negative Independent Component Analysis problem (NNICA). NNICA original separation criterion contains the discontinuous *sign* function whose minimization may lead to ill convergence (local minima) especially for sparse sources. Replacing the discontinuous function by a continuous one *tanh*, we propose a more accurate regularized Gradient algorithm called “Exact” Regularized Gradient (ERG) for NNICA. Experiments on synthetic data with different sparsity degrees illustrate the efficiency of the proposed method and a comparison shows that the proposed ERG outperforms existing methods.

Index Terms— Non-negativity, Independent Components Analysis, Gradient descent, Sparsity, Convergence Algorithms, Well-grounded sources

1. INTRODUCTION

Independent Component Analysis is a well-known technique for solving blind source separation problem [1][2]. Restricted to the noiseless linear instantaneous square mixture, the ICA model is given by:

$$X = AS \quad (1)$$

$n \times p \quad n \times nn \times p$

where n is the number of sources and p the number of samples. S are the n hidden independent sources, A is the unknown mixing matrix and X are the observations.

ICA aims at estimating the sources S and the mixing matrix A when only the observations X are given.

Many effective ICA algorithms [1][3][4][5] were proposed for solving (1). However “classical” ICA algorithms seem not suited for non-negative sources. Under non-negativity constraints on S and A , Non-negative Matrix Factorization (NMF) [6][7][8] was proposed for solving (1) where the estimated sources and the mixing matrix are all constrained to be non-negative. But NMF techniques still suffer from initialization problem and the non-negativity alone is not sufficient to guarantee uniqueness of the solution as shown in [9]. Another approach uses a Bayesian method [10] for solving (1) under the non-negativity constraint of S and/or A , but this approach can be

computationally complex and time-consuming especially for large scale problems. By taking into account both non-negativity and mutual independence, Non-Negative Independent Component Analysis (NNICA) has been introduced for solving (1) under non-negativity constraint on S , A being positive or of mixed sign. NNICA was first introduced for independent and “well-grounded” sources [11]. Zheng et al. proposed an extension to non well-grounded sources using Neural Network and the minimization of the mutual information [12].

This paper considers the NNICA problem for well-grounded sources. In the next section we summarize the NNICA problem and list the main approaches used for solving it. Section 3 describes the proposed “Exact” Regularized Gradient (ERG) algorithm. Since ERG will be applied to real mass spectra data, simulations are performed on synthetic data with different sparsity degrees and results are presented in section 4. In fact, mass spectra are sparse data then verifying the well-grounded constraint. One may note that sparse data are more sensitive to ill convergence problem due to discontinuity. We also compare ERG to existing NNICA methods. Section 5 presents the conclusions and future works.

2. NNICA PROBLEM AND EXISTING APPROACH

Let consider the model given by (1) and suppose that the hidden sources $S = [s_1 s_2 \dots s_n]^T$ are *non-negative* ($Pr(s_i < 0) = 0, \forall 1 \leq i \leq n$), *independent* ($Pr(s_1, s_2, \dots, s_n) = \prod Pr(s_i)$), and *well grounded* ($\forall \delta > 0, Pr(s_i < \delta) > 0, \forall 1 \leq i \leq n$).

The unknown mixing matrix A being assumed to be positive or of mixed sign, the aim of NNICA is to estimate the sources given only the observations X .

Under the previous assumptions, it has been shown that the sources can be correctly estimated by first whitening the observations and second rotating the whitened data to fit them on the positive orthant [11]. Let $Z = VX$ be the whitened observations and the output $Y = WZ$ be the rotated data (V being a whitening matrix and W a rotating one).

The performance index to minimize is blindly defined as [13]:

$$J_0(W) = \frac{1}{2} \|Z - W^T Y^+\|_F^2 = \frac{1}{2} \|Y^-\|_F^2 \quad (2)$$

where $[Y^+]_{ij} = \max(0, Y_{ij})$ and $[Y^-]_{ij} = \min(0, Y_{ij})$. NNICA reduces to solving the following optimization problem:

$$W^* = \arg \min_{W \in SO(n)} J_0(W) \quad (3)$$

where $SO(n)$ is the special orthogonal group (the group of orthogonal matrices of determinant 1).

This study is conducted in the context of co-tutelle Ph-D program between CEA List, GIPSA Lab of University of Grenoble (France) and Signals and Systems Laboratory of National School of Engineers of Tunis (Tunisia).

The main approaches for solving (3) include Non-negative PCA (NNPCA) [14][15], Axis Pair Rotation (APR) [13][16], Geodesic search (GEO) [17] and Projected Gradient (PG) [18].

3. “EXACT” REGULARIZED GRADIENT METHOD FOR NNICA

To deal with the optimization problem (3) in one step (and thus avoiding a “projection” step), we slightly modify the original criterion J_0 by adding a penalty term which penalizes the deviation to orthogonality [15]. The new criterion is given by :

$$J(W) = \frac{1}{2} \|Y^-\|_F^2 + \gamma \|W^T W - I_n\|_F^2 \quad (4)$$

and the optimization problem to solve becomes:

$$W^* = \arg \min_{W \in \mathcal{M}_n} J(W) \quad (5)$$

where \mathcal{M}_n is the set of square dimension n matrix.

For solving (4), we propose to use the gradient descent defined by the iterative algorithm:

$$W^{k+1} = W^k - \mu [\nabla_W J]_{W=W^k} \quad (6)$$

But when developing the criterion, one may note that (due to $\|Y^-\|_F^2$ term) it contains the discontinuous function *sign* as shown in the following expression:

$$J(W) = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^p \left(\frac{Y_{ij}}{2} (1 - \text{sign}(Y_{ij})) \right)^2 + \gamma \sum_{k=1}^n \sum_{l=1}^n \left(\sum_{t=1}^n W_{tk} W_{tl} - \delta_{kl} \right)^2 \quad (7)$$

When computing the gradient, this leads to Dirac distributions which are neglected in [18] giving an “approximate gradient expression”. This approximation may lead to ill convergence (local minima) in the context of blind source separation because of the neglected terms. For a more accurate expression, we propose to approximate the discontinuous function *sign* by a continuous one *tanh* for “exact calculation”. We then introduce the criterion J_λ given by:

$$J_\lambda(W) = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^p \left(\frac{Y_{ij}}{2} (1 - \tanh(\lambda Y_{ij})) \right)^2 + \gamma \sum_{k=1}^n \sum_{l=1}^n \left(\sum_{t=1}^n W_{tk} W_{tl} - \delta_{kl} \right)^2 \quad (8)$$

where the parameter λ controls the accuracy of the *sign* function approximation: the larger is λ , the better is the approximation since one may note that $J_\lambda \rightarrow J$ as $\lambda \rightarrow +\infty$. So the gradient expression is given by:

$$[\nabla_W J_\lambda]_{ij} = \frac{1}{2} \sum_{r=1}^p \frac{Y_{ir} (1 - \tanh(\lambda Y_{ir}))}{2} (1 + \beta) Z_{jr} + 4\gamma \sum_{l=1}^n W_{il} \left(\sum_{t=1}^n W_{tj} W_{tl} - \delta_{ij} \right) \quad (9)$$

where $\beta = -\tanh(\lambda Y_{ir}) - \lambda Y_{ir} [1 - \tanh^2(\lambda Y_{ir})]$. The obtained iterative algorithm is given by :

$$W_{ij}^{k+1} = W_{ij}^k - \mu [\nabla_W J_\lambda]_{ij} \quad (10)$$

where μ is a small positive update step.

Comparing to gradient expression computed from J_0 criterion in [18], if we replace *sign* by *tanh* we get :

$$[\nabla_W J_0]_{ij}^\lambda = \frac{1}{2} \sum_{r=1}^p \frac{Y_{ir} (1 - \tanh(\lambda Y_{ir}))}{2} (1 + 1) Z_{jr} \quad (11)$$

We note that the term β in the first right-side part of equation (9) is replaced by 1 in equation (11). One may expect better convergence of algorithm (10) because when converging, the estimated sources become well-grounded (since the original sources are assumed to be well-grounded) and the term β becomes non negligible as illustrated on Figure 1.

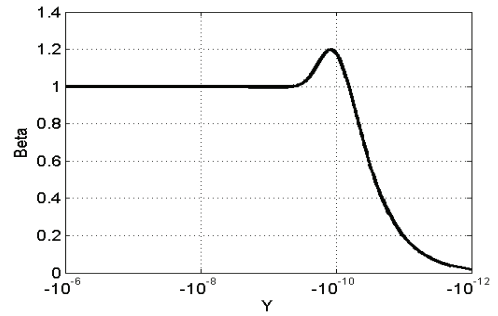


Fig. 1. β term vs Y value for $\lambda = 10^{10}$

4. SIMULATION RESULTS

The proposed algorithm has been evaluated on synthetic data and compared to existing ones.

The non-negative sources matrix has been generated by the *Matlab* “sparse uniformly distributed random matrix generator” (*sprand* function) which uses three parameters: the number of rows (number of sources n), the number of columns (number p of samples for each source) and the sparsity degree *spar*, which determines the non-zero element ratio in the source S matrix ($0 \leq \text{spar} \leq 1$).

The mixing matrix A has been generated using the *Matlab* “normally distributed random generator” (*randn* function).

Three performance measures are used for evaluation:

- The reconstruction error defined by the equation (12). It is a blind performance index measuring the negativeness of the output, it corresponds to the original criterion to minimize.

$$E_{rec} = \frac{1}{np} \|Y^-\|_F^2 \quad (12)$$

- The separation error defined by the equation (13). It is a non blind performance index measuring the separation quality, it is similar to an inter-symbol interference ratio.

$$E_{sep} = \frac{1}{n(n-1)} \left[\sum_i \left(\sum_j \frac{|(WVA)_{ij}|}{\max_l |(WVA)_{il}|} - 1 \right) \right] + \frac{1}{n(n-1)} \left[\sum_j \left(\sum_i \frac{|(WVA)_{ij}|}{\max_l |(WVA)_{lj}|} - 1 \right) \right] \quad (13)$$

- The CPU time evaluates the time to convergence using Intel(R) Core(TM) 2 Duo CPU P8400 computer.

In all the simulations, we set the source number $n = 10$ and the sample number $p = 1000$. Three different sparsity degrees were considered:

- case 1: $spar = 1$ corresponding to 100% of non-zero entries in the source matrix.
- case 2: $spar = 0.1$ corresponding to 10% of non-zero entries in the source matrix.
- case 3: $spar = 0.01$ corresponding to 1% of non-zero entries in the source matrix, this situation corresponds to our future condition for mass spectra data.

A theoretical analysis of the influence of the parameter λ will be consider later, nevertheless as mentionned in section 3, λ must be setted to large value for good approximation of the *sign* function by the tanh one.

The algorithm parameters λ , μ and γ were experimentally tuned for optimal convergence according to the separation error. We set $\mu = 0.5$, $\gamma = \frac{1}{16}$, $\lambda = 10^{10}$ in all the simulations and W^0 initial value W^0 was setted to identity matrix ($W^0 = I_n$).

Figure 2 shows the average performance index obtained with 50 Monte Carlo runs for the different sparsity degrees. We can see that the proposed ERG algorithm seems slightly slower than others method but it always presents the smallest reconstruction and separation errors. The same results (not shown) are observed when modifying the number of sources n and/or the number of samples p .

Figure 3 shows the ERG separation ability, we can see that the sources are successfully recovered.

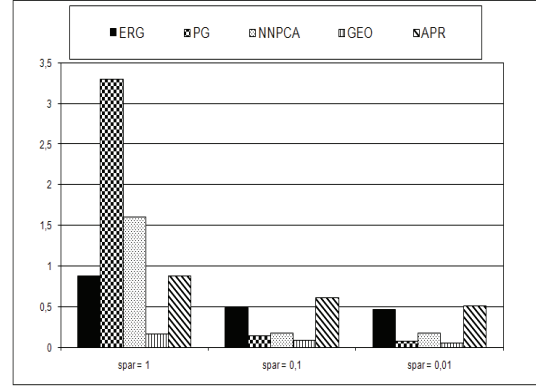
Figure 4 shows the reconstruction error E_{rec} and the separation error E_{sep} versus the iterations number for one run of the case 3. This figure shows that the proposed ERG converge to better solution than other methods.

5. CONCLUSIONS AND FUTURE WORKS

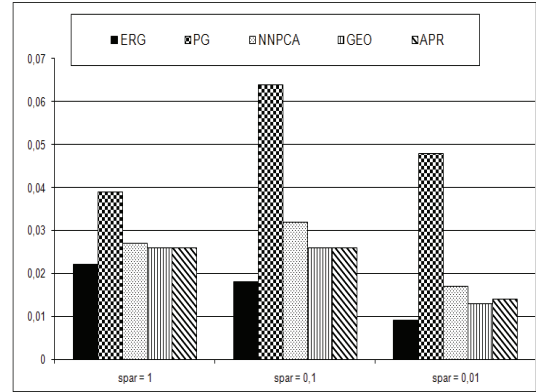
This paper considers the Non-negative Independent Component Analysis problem for well-grounded sources. By slightly modifying the separation criterion, we rewrite the original constrained optimization problem as a non constrained one. For more accurate calculation, we approximate the discontinuous function *sign* used in the criterion by a continuous one tanh. This lead to an “Exact” Regularized Gradient algorithm improving the algorithm convergence. Simulations on synthetic data with different sparsity degrees highlight the fact that proposed ERG outperforms existing methods. For future work, a theoretical convergence analysis will be considered to determine the algorithm optimal parameters. Incorporating sparsity “a priori” in the method can be an interesting approach since ERG will be applied to real mass spectra data. Considering noisy mixtures can also be a good challenge for evaluating the algorithm robustness. Finally a quasi Newton approach can be investigated in order to develop a faster algorithm.

6. REFERENCES

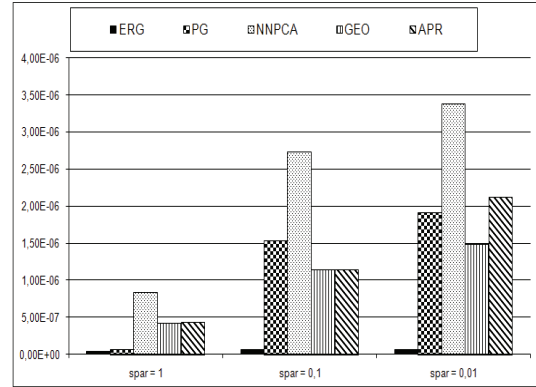
- [1] P. Comon and C. Jutten, Eds., *Handbook of Blind Source Separation: Independent Component Analysis and Applications*, Academic Press, 2010.



(a) Average CPU time (s)



(b) Average separation error



(c) Average reconstruction error

Fig. 2. Comparaion of the average performance index vs sparsity degree for the algorithm ERG, PG, NNPCA, GEO and APR ($n = 10$ and $p = 1000$)

- [2] P. Comon, “Independent Component Analysis, A new concept?,” *Signal Processing*, vol. 36, pp. 287–314, 1994.
- [3] J. F. Cardoso and A. Souloumiac, “Blind Beamforming for Non Gaussian Signal,” *IEE Proceedings-F*, vol. 140, pp. 362–370, 1993.
- [4] A. Hyvärinen, “Fast and Robust Fixed-Point Algorithms for Independent Component Analysis,” *IEEE Transactions on Neural Networks*, vol. 10, no. 3, pp. 626–634, 1999.

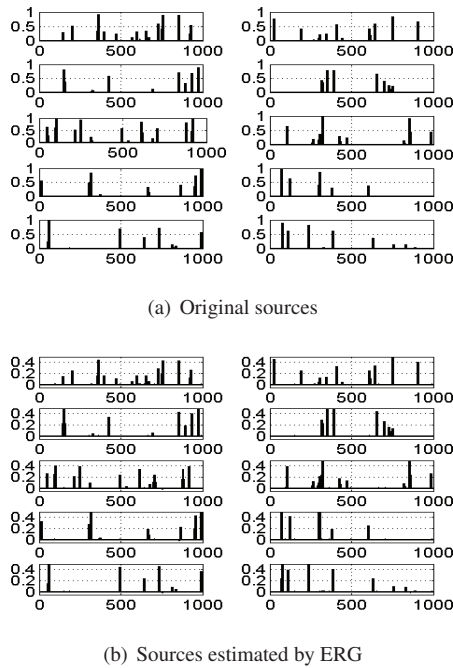


Fig. 3. ERG separation ability ($n = 10$, $p = 1000$, $spar = 0.01$)

- [5] J. F. Cardoso, E. Moulines, A. Belouchrani, and K. A. Meraim, "A Blind Source Separation Technique Using Second-Order Statistics," *IEEE Transactions on Signal Processing*, vol. 45, pp. 434–444, 1997.
- [6] D. D. Lee and H. S. Seung, "Algorithms for Non-negative Matrix Factorization," *Neural Information Processing Systems*, vol. 13, pp. 556–562, 2001.
- [7] C. J. Lin, "Projected Gradient Methods for Non-negative Matrix Factorization," *Neural Computation*, vol. 19, no. 10, pp. 2756–2779, 2007.
- [8] A. Cichocki, R. Zdunek, A. H. Phan, and S. Amari, *Nonnegative Matrix and Tensor Factorizations, Applications to Exploratory Multi-way Data Analysis and Blind Source Separation*, John Wiley and Sons, Ltd, 2009.
- [9] S. Moussaoui, D. Brie, and J. Idier, "Non-negative Source Separation: Range of Admissible Solutions and Conditions for the Uniqueness of the Solutions," *IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 5, pp. 289–292, 2005.
- [10] S. Moussaoui, D. Brie, A. M. Djafari, and C. Carteret, "Separation of Non-Negative Mixture of Non-Negative Sources Using a Bayesian Approach and MCMC Sampling," *IEEE Transactions on Signal Processing*, vol. 54, no. 11, pp. 4133–4145, 2006.
- [11] M. Plumbley, "Conditions for Nonnegative Independent Component Analysis," *IEEE Signal Processing Letters*, vol. 9, pp. 177–180, 2002.
- [12] C. H. Zheng, D. S. Huang, Z. L. Sun, M. R. Lyu, and T. M. Lok, "Nonnegative independent component analysis based on minimizing mutual information technique," *Neurocomputing*, vol. 69, pp. 878–883, 2006.

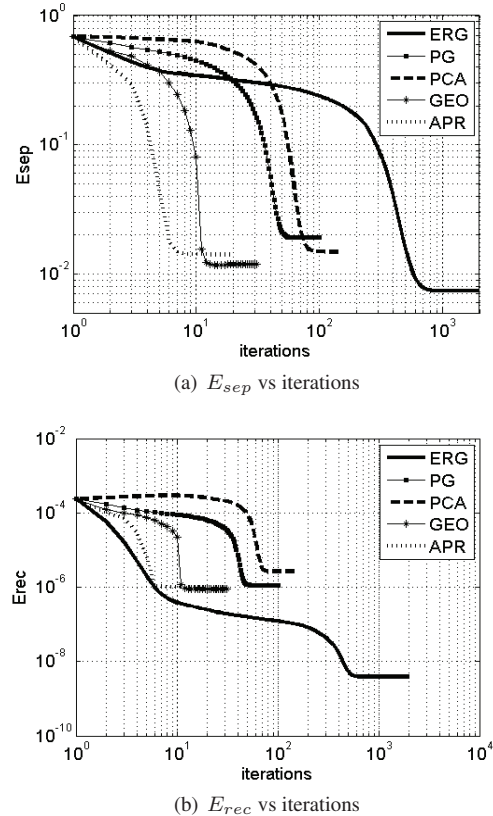


Fig. 4. Performance vs iteration number ($n = 10$, $p = 1000$, $spar = 0.01$)

- [13] M. Plumbley, "Algorithms for Nonnegative Independent Component Analysis," *IEEE Transactions on Neural Networks*, vol. 14, no. 3, pp. 534–543, 2003.
- [14] M. Plumbley, "A Nonnegative PCA Algorithm for Independent Component Analysis," *IEEE Transactions on Neural Networks*, vol. 15, no. 1, 2004.
- [15] Y. Li and H. Zheng, "Improvement for Nonnegative PCA Algorithm for Independent Component Analysis," *Neural Networks and Brain, 2005. ICNN&B '05. International Conference on*, vol. 3, pp. 2000–2002, 2005.
- [16] W. S. B. Ouedraogo, A. Souloumiac, and C. Jutten, "Non-negative Independent Component Analysis Algorithm Based on 2D Givens Rotations and a Newton Optimization," *Lecture Notes in Computer Science*, vol. 6365, pp. 522–529, 2010.
- [17] M. Plumbley, "Optimization using Fourier Expansion over a Geodesic for Non-Negative ICA," *Lecture Notes in Computer Science*, vol. 3195, pp. 44–56, 2004.
- [18] E. Oja and M. Plumbley, "Blind Separation of Positive Sources by Globally Convergent Gradient Search," *Neural Computation*, vol. 16, no. 9, pp. 1811–1825, 2004.